

A homology model for *Clostridium difficile* methionyl tRNA synthetase: active site analysis and docking interactions

Ehab Al-Moubarak · Claire Simons

Received: 11 August 2010 / Accepted: 5 October 2010 / Published online: 2 November 2010
© Springer-Verlag 2010

Abstract Treatment of *C. difficile* infection is one of the most difficult biomedical challenges. To develop novel antibacterials, researchers have been targeting bacterial molecular functions that are essential for its growth. The methionyl tRNA synthetase (MetRS) is strictly required for protein biosynthesis and success was reported in developing antibacterials to inhibit this enzyme. The present study was aimed at building and analyzing a homology model for *C. difficile* MetRS in the context of drug design. A homology model of *C. difficile* MetRS was constructed using Molecular Operating Environment (MOE) software. *A. aeolicus* MetRS was the main template while the query zinc binding domain was modeled using *T. thermophilus* MetRS. The model has been assessed and compared to its main template (Ramachandran, ERRAT and ProSA). The active site of the query protein has been predicted from its sequence using a detailed conservation analysis (ClustalW2). Using MOE software, suitable ligands were docked in the constructed model, including a *C. difficile* MetRS inhibitor REP3123 and the enzyme natural substrate, and the key active site residues and interactions were identified. These docking studies have validated the active site conformation in the constructed model and identified binding interactions.

Keywords *Clostridium difficile* · Homology model · Methionyl tRNA synthetase · Molecular docking · MOE

Introduction

Clostridium difficile (*C. difficile*) is the most common cause of healthcare related diarrhoea in hospitals and health care institutions. Hospitalized patients on broad spectrum antibiotic treatment are considered those who are particularly at risk from developing the disease, an example is patients suffering from spinal cord injury who are subject to frequent antibiotic exposure and are hospitalized or using hospital facilities for long periods [1]. It is estimated there has been a 40 percent increase in the *C. difficile* infection (CDI) incidence from 2006 to 2008 [2] and news sources have shown that in 2008 8,324 people in England and Wales died while affected by the disease, with CDI the cause of death in 50% of cases [3]. This figure shows a four times increase in the deaths associated with CDI compared with 2001 [3].

C. difficile infection is currently treated with a range of therapies with meronidazole, routinely used as initial therapy for CDI [2]. Discontinuation of antibiotic treatment can restore the gut normal flora and stop the *C. difficile* overgrowth that results in the disease incident [2]. Nevertheless this is not always feasible especially when the patient has been treated for a serious condition, moreover, in the best situations only about 20–25 percent benefit from this intervention. Vancomycin is used in more severe cases and in subsequent episodes; combinations of antimicrobials

E. Al-Moubarak · C. Simons (✉)
Medicinal Chemistry, Welsh School of Pharmacy,
Cardiff University,
King Edward VII Avenue,
Cardiff CF10 3NB, UK
e-mail: simonsc@cardiff.ac.uk

have also been tried in some cases and treatment guidelines and algorithms are available to help in the management of CDI cases [2].

A potential target in the development of drugs for the treatment of *C. difficile* infections is aminoacyl-tRNA synthetase (AaRS). Aminoacyl-tRNA synthetases, can exist as either homodimers or monomers and catalyze the specific aminoacylation of amino acid transfer RNAs [4]. Class I aminoacyl synthetases consist of four domains: a catalytic Rossmann fold, a connective peptide (CP) inserted between the two halves of the Rossmann fold, a KMSK domain and a C-terminal α -helix bundle domain, responsible for the recognition of the CAU anticodon of aminoacyl tRNAs [5], whereas class II synthetases are characterized by a catalytic core of antiparallel β -sheets and three consensus motifs [6].

Although the natural substrates for AaRS enzymes are the same for all living organisms, the mode of binding of these amino acid substrates can be different depending on several factors including the active site conformation. In recent years increasing numbers of compounds have been reported to inhibit AaRS enzymes, many of them however inhibit human enzymes [7], which can cause toxicological issues. Success in identifying molecules that specifically inhibit bacterial AaRS enzymes has recently been achieved [7]. Success in clinical applications started with Mupirocin (marketed as Bactroban), which is the first AaRS inhibitor antibiotic that selectively inhibits bacterial isoleucyl-tRNA synthetase [7, 8]. Mupirocin is now the first choice topical antibiotic for the treatment of methicillin-resistant *Staphylococcus aureus* (MRSA).

Several potent nanomolar *S. aureus* methionyl tRNA synthetase (MetRS) inhibitors were developed and found to have excellent antibacterial activity against *Staphylococcal* and *Enterococcal* pathogens [9–12]. One of the quinolinone derivatives tested selectively inhibited *S. aureus* MetRS and was found to be effective *in vivo* in rat [9]. Very interesting is the generation of REP3123, which is a potent MetRS inhibitor of *C. difficile* [13]. REP3123 was developed as an anti-microbial agent for treatment of *C. difficile* associated diseases (CDAD) and showed potent antimicrobial activity against *C. difficile* and Gram-positive aerobic bacteria; however it does not affect the normal gut flora content. REP3123 prevents *C. difficile* sporulation and therefore inhibits toxin production. The compound is an effective inhibitor of both *C. difficile* and *S. aureus* MetRS and does not inhibit human MetRS [13]. The success of REP3123 as an inhibitor of *C. difficile* MetRS provides a useful ligand, in addition to known substrates of MetRS, to develop and evaluate a homology model.

There are no crystal structures of the AaRS enzymes from *C. difficile*, however comparative structure modeling can be used to construct a putative structure for the desired

C. difficile MetRS enzyme that may facilitate inhibitor development. The construction and subsequent evaluation of a *C. difficile* MetRS homology model is described.

Methods

Construction of the *C. difficile* MetRS model

Homology searching

The *C. difficile* MetRS sequence was obtained from the ExPASy proteomics server [14]. The enzyme sequence has the Uniprot identifier Q181D9 and was obtained from *C. difficile* strain 630, it has 645 residues [15]. A homology search was performed using NCBI BLAST2 service [16] accessible from the ExPASy server, which was used to align the query against sequences in the Protein Data Bank (PDB) [17] and thus the close homologous proteins were identified. The alignment parameters and thresholds used for screening candidate homologues were used with their default values and BLOSUM62 comparison matrix.

The Phylogeny server [18, 19] was used to build a phylogenetic tree for these identified homologous proteins, the query sequence and other MetRS enzymes selected from different organisms.

Multiple sequence and structure alignment

The secondary structure of *C. difficile* MetRS together with the templates were determined using PSIPRED [20]. The query sequence was then aligned against the most homologous templates, *Aquifex aeolicus* MetRS, *Thermus thermophilus* MetRS, *Pyrococcus Abyssii* MetRS, *Escherichia Coli* MetRS using Clustal W2 [21] for identification of specific α -helices, β -sheets, coils and loops, and the common features and motifs: 'HIGH region', KMSKS motifs, zinc metal and ATP binding sites. Further evaluation of fold recognition and identification of template structures by structural alignment was performed using mGenTHREADER [22].

3D model building

The molecular modeling experiments were performed on a RM Intel Core2 Duo CPU 2.2 GHz running Windows XP using Molecular Operating Environment (MOE) 2008.09 molecular modeling software [23]. Homology models were built using MOE-Homology using AMBER99 forcefield [24], which uses a dictionary to set the partial charges of atoms in amino acids. The final homology model was constructed using all the *A. aeolicus* MetRS (2CSX) crystal structure and the *T. thermophilus* MetRS zinc finger

Table 1 The five optimal crystal templates identified in the *C. difficile* MetRS BLAST results

Enzyme in organism	PDB code	Blast score ^a	Sequence identity ^b	Seq. identity %	Chain length	E. value
Aquifex aeolicus	2CSX-A	449	232/480	48%	479	e-127
Thermus thermophilus	1A8H-A	363	201/509	39%	500	e-101
Pyrococcus Abyssii	1RQG-A	290	206/719	28%	722	7e-79
Escherichia Coli	1F4L-A	188	131/553	23%	551	5e-48
C terminal of Pyrococcus Abyssii	1MKH-A	100	50/92	54%	107	1e-21

All were MetRS enzymes

^a The BLAST score for an alignment is calculated by summing the scores for each aligned position and the scores for gaps

^b (Number of identical residues)/ (length of sequence fragment identified by BLAST)

residues (from 1A8H). MOE-align was used to identify the zinc binding domain residues in the second template (*T. thermophilus* MetRS). The identified residues were selected to override the main template for building the model, the zinc and the Met-tRNA atoms were selected and included in the model as environment for induced fit, all other setting options were set to the default. Ten intermediate models were generated and the final model was taken as the Cartesian average of all the intermediate models. All the minimizations were performed with MOE until RMSD gradient of $0.05 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ with the forcefield specified and the partial charges were automatically calculated.

Model validation

Stereochemical quality of the polypeptide backbone and side chains was evaluated using Ramachandran plots obtained from the RAMPAGE server [25]. Amino acid environment was evaluated using ERRAT plots [26]. ERRAT assesses the distribution of different types of atoms with respect to one another in the protein model. The ProSA tool [27] was used to check the overall model protein structure for potential errors. Validation data for the templates *A. aeolicus* MetRS (2CSX) and *T. thermophilus* MetRS (1A8H) was used as the baseline to assess the respective models.

Docking

Ligand structures were obtained from the relevant complex crystal structures, REP3123 was built using MOE-builder

[23] and then the ligand was energy minimized. These ligands were docked using the MOE-Dock with setting: Placement: Triangle Matcher; Rescoring 1: London dG, 30 poses were constructed for each compound and the best scoring model-ligand complexes were selected; the ligand interactions within these complexes were visualized using the MOE ligand interactions simulation.

Results and discussion

Homology models and validation

Initial screening for possible templates was performed by running a BLAST analysis [16] of the amino acid sequence of *C. difficile* MetRS, obtained from the ExPASy proteomics server [14], against the PDB resolved structures, identified five structures for further consideration as possible templates (Table 1). For a structure to be considered it should ideally be wild-type, rather than mutant or engineered, have reasonable sequence identity with the *C. difficile* MetRS and preferably have the same function. The first ten native hits are bacterial MetRS enzymes, with the MetRS enzyme of *Aquifex aeolicus* (*A. aeolicus*) being the most appropriate template due to the high sequence identity (48%). The *Thermus thermophilus* (*T. thermophilus*) enzyme also had a good sequence identity of 39%. In addition *Escherichia Coli* (*E. coli*) MetRS (23% identity) and *Pyrococcus Abyssii* (*P. Abyssii*) MetRS (28% identity) can still be considered to have a reasonable sequence similarity. These MetRS enzymes constitute the dataset of suggested templates.

Table 2 Key residues of the suggested templates

PDB	HIGH region motif	KMSKS motif	Zinc binding residues	ATP binding
2CSX-A	14-18	295-299	129, 132, 147, 150	298
1A8H-A	12-22	297-301	127, 130, 144, 147	300
1RQG-A	11-21	344-348	143, 146, 156, 159	347
1F4L-A	15-25	333-337	146, 149, 159, 162	336

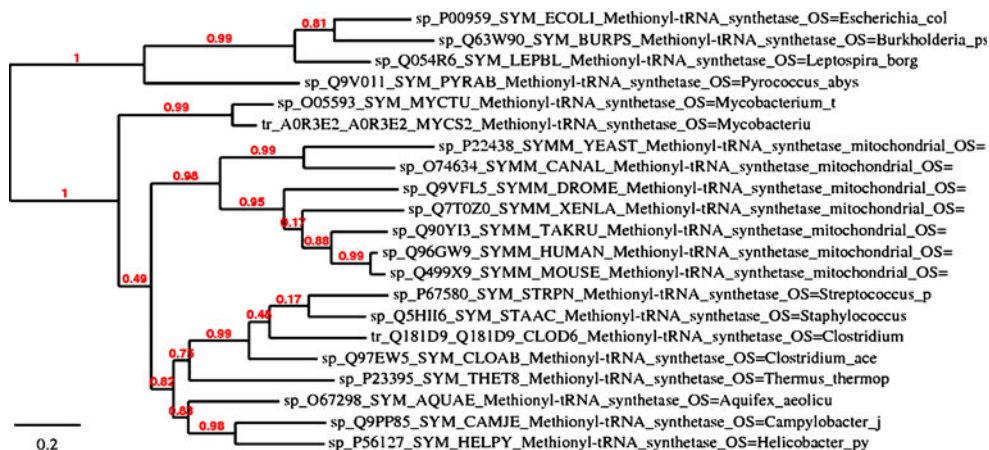


Fig. 1 The phylogenetic tree of *C. difficile* MetRS in relation to MetRS enzyme from other organisms. *Escherichia Coli* (P00959); *Pseudomonas pseudomallei* (Q63W90); *Bovine* (Q054R6); *Pyrococcus Abyssii* (Q9V011); *Mycobacterium tuberculosis* (O05593); *Mycobacterium smegmatis* (A0R3E2); *Saccharomyces cerevisiae* (*Baker's yeast*) (P22438); *Candida albicans* (O74634); *Drosophila metanogaster* (*Fruit fly*) (Q9VFL5); *African clawed frog* (Q7T0Z0);

Japanese pufferfish (Q90YI3); *Human* (Q96GW9); *Mouse* (Q499X9); *Streptococcus pneumoniae* (P67580); *Staphylococcus aureus* (Q5H116); *Clostridium difficile* (Q181D9); *Clostridium acetobutylicum* (Q97EW5); *Thermus thermophilus* (P23395); *Aquifex aeolicus* (O67298); *Campylobacter jejuni* (Q9PP85); *Helicobacter pylori* (P56127)

The MetRS enzymes have common features and motifs: 'HIGH region', KMSKS motif, zinc metal and ATP binding sites (Table 2). Class I AaRS enzymes are characterized by a catalytic center built around a Rossmann fold where the two signature sequences HIGH and KMSKS are found [5, 28]. The HIGH motif is located in a loop between $\beta 1$ and $\alpha 1$ elements, with the two histidine residues stacked upon each other [28]. The KMSKS peptide is located in a β - α - α - β - α structure, named the stem contact (SC) fold [29, 30]. The HIGH region has been implicated in amino acid activation and the KMSKS in docking the acceptor stem of tRNA. The positions of these features in different species are varied or shifted by a few residues but are still present in all of them [5].

To obtain more information regarding the best possible template, the Phylogeny server [18, 19] was used to construct a phylogenetic tree using sixteen more MetRS enzyme sequences in addition to the query and the above four suggested MetRS templates. A diverse range of species was considered including mammals, fish, yeast and bacteria so as to appreciate the relative distances between the suggested four templates and the query sequence (Fig. 1).

The enzyme sequences of *S. aureus* and *C. acetobutylicum* were the closest to the query sequence in the phylogenetic tree. However, by considering the next closest branch, it was interesting to observe that the enzyme sequences of *A. aeolicus* and *T. thermophilus* were the next closest. The other two suggested templates of *E. coli* and *P. Abyssii* were diverting in branches relatively far from the query sequence.

Multiple sequence and structural alignments

ClustalW2 [21] was used to align the suggested template sequences and the query sequence of *C. difficile* MetRS (Fig. 2).

Very few gaps in alignment were observed when considering the *C. difficile*, *T. thermophilus* and *A. aeolicus* MetRS sequences alone until approximately residue 500. However, consistent gaps were observed between these three sequences in respect to the other two suggested templates of *E. coli* and *Pyrococcus Abyssii* MetRS sequences. The difference between the two groups was in agreement with the relations described by the phylogenetic tree (Fig. 1). HIGH and KMSKS motifs are identified in all suggested templates and query sequence (Fig. 2).

To compare structural alignment, the *C. difficile* MetRS protein sequence was used as a template for the mGenTHREADER program [22], to search the Protein DataBank (PDB) for proteins that are likely to contain similar structural motifs. The mGenTHREADER program is a recommended method for fold recognition and identification of distant homologues and makes use of profile-profile alignments and predicted secondary structure (using

Fig. 2 Sequence alignment of MetRS enzymes of Aquifex aeolicus, Clostridium difficile, Thermus thermophilus, Pyrococcus Abyssii and Escherichia Coli (E) using Clustal W2 in which "*" means that the residues are identical, ":" means that conserved substitutions have been observed, "." means that semi-conserved substitutions are observed. The residues are colored according to their chemical properties where red, small hydrophobic (AVFPMILWY); blue, acidic (DE); purple, basic (RHK); green, hydroxyl + amine + basic (STYHCNGQ)

CLUSTAL 2.0.12 multiple sequence alignment

```

Aquifex      MTLMKKFYVITPIYVNDVHLGHAY-TTIAADTIARYYRLRDYDVFFLTGTDEHGLKIQ 59
Clostridium  -MSKPSFYVITPIYYPSSGGLHIGHTY-STVAADTIARFKRFCGYDVKFLTGTDEHGEKIQ 58
Thermus      --MEKVIFYVITPIYVNAEHLGHAY-TTVVADFLARWHRLDGYRTFFLTGTDEHGETVY 57
pyrococcus   ---MVRVMVTSALPYANGPIHAGHLAGAYLPADIFVRYLRKLGEDVVFICGTDEHGTPIIS 57
E            ---AKKILVTCALPYANGSIHLGHML-EHIQADVVRVYQRMRGHEVNFICADDAHGTPIM 56
          ** .: * . * ** : ** .*: * : . . * : . * ** :

Aquifex      KKAEEELGISPKELVDRNAERFKKLWVFLKIEYTKFIRTTDPYHVKVFQKVFEECYKRGDI 119
Clostridium  KKAIEQGMSEIEYLDGMIKDIKALWNTMDSYDDFIRTTTEKRHTDIIQKIFTKLYEQGDI 118
Thermus      RAAQAAGEDPKAFVDRVSGRFKRAWDLGLIAYDDFIRTTTEERHKKVQVLVKKVYEAGDI 117
pyrococcus   FRALKEGRSPREIVDFEHEQIKITFQRAKISDFDFGRTELPPIHYKLSQFKLWAYENHGL 117
E            LKAQQLGITPEQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYSRLKENGFI 116
          * * : . : : * : : * : . . : . : * :

Aquifex      YLGEYEGWY-----CVGCEEFKSEAEALAEHDHTCPPIH 150
Clostridium  YKGEYEGRY-----CTFCESFWTESQLLEGNKCPDC 149
Thermus      YYGEYEGLY-----CVSCERFYTEKELVEG-ICPIH 147
pyrococcus   VKKVTKQAYCEHDKMFLPDRFVIGTCPCYGAEDQKGDQCEVCGRPLTPEILLINP-RCAIC 176
E            KNRTISQLYDPEKGMFLPDRFVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEP-KSWVS 175
          . * : * * : * * : * * : * * :

Aquifex      QKKCEYIKEPSYFFRLSKYQDKLLELYEKNPEFIQPDYRRNEIIS-FVKQGLKDLVSRP 209
Clostridium  GRETYLVKEESYFFRLSKYEDRLKELFKDDS-FCFPAARKNEMVANFLDKGLEDLVSRP 207
Thermus      GRPVERRKEGNVFFRMEKYRPLWQEQYIQENPDLIRPEGYRNEVLA-MLAEPIDLSISR 206
pyrococcus   GRPISFRDSAHHYIKMQDFAERLKRWIEKQP---WKPNVKNMVLVSWIEGLEERAITR 231
E            GATFVMRDESHFFDLPSFSEMLQAWTRSGA---LQEQVANKMQEWFESGLQQWDISR 230
          . . : : : : * . . . : : . . : : : **

Aquifex      RSRVKWGI PVFPD---PEHTIYVWFDALFNYSAL-----EDKVEIYWP----- 250
Clostridium  -TTFDWGIKVPFD---EKHVIYVWFDALCNYITALGYMTD---NDEEFKKYWP----- 253
Thermus      KSRVPWGIPLPFD---ENHVIYVWFDALLNYVSALDYP---EGEAYRTFWP----- 251
pyrococcus   --DLNWGI PVPLDEEDMKGVLYVWFAPIGYISITIEHFKRIGKPNWKKYWLNDGQT 289
E            --DAPY---FGFEIPNAPGKYFYVWLDAPIGYMGSFKNLCKRGRDVSFDEYWK-KDSTA 284
          : . : : : * ** : * : : * :

Aquifex      ADLHLVGDILRHHIVYVWPAFLMSLG-----YELPKKVFAGHWVTEG-KKMSKT 299
Clostridium  ANVQIVGKEIVRHHIIVWPAFLMALG-----LEVPKQVFGHWILFAD-DKMSKS 302
Thermus      HAWHLIGKDILKHHAVFWPTMLKAAG-----IPMYRHLNVGGFLGPDGRKMSKT 301
pyrococcus   RVIHFIGKDNIPHHAFVWPAFLMAYGKYKDEEVEAEWNLVYDIPANEYLTLEG-KKFSKS 348
E            ELYHFIGKDIVYHSLFWPAMLEGS-----NFRKPSNLFVHGYVTVNG-AKMSKS 333
          : : : : * : * : * : . : : : . * : :

Aquifex      LGNVDPYEVVQYGLDEVRYFLLREVP-FGQDGFSSKAILNRINGELANEIGNLYSRV 358
Clostridium  KGNVVYPEPIIERYGIDTLKYFLLREFA-FGQDGSYTHRNFTVTRINVDLANDLGNLISRT 361
Thermus      LGNVVDPFALLEKYGRDALRYLLREIP-YGQDTPVSEALRTRYEADLADDLGNLVQRT 360
pyrococcus   RNWAIWVHEFLDVFPADYLRYYLTIIMP-ETRDSDFSFDFKVRINEELVNNLGNVHRA 407
E            RGTFIKASTWLNHFADSLRYYYTAKLSSRIDDI DLNLEDFVQRVNADIVNKKVNLASRN 393
          . : : : * : * : . . * . : * : : : : * : *

Aquifex      VNM AHKFLG----GEVSGARDEEYAKIAQESIKNYENYMEKVNFKYKAIIEEILKFTSYLN 413
Clostridium  VAMVEKYNNGIIPITAKVSTDFDADLKEQAVSTRENFEAEMDKMQFHEALESVWKLVRRTN 421
Thermus      RAMLFRFAEG----RIPEPVAGEELAEGLAGRLRPLVRELKHFHVALEAMAYVKALN 415
pyrococcus   LTFVNRVYFDGVVPERGELDELDRALLEEIEKAFKEVGEIMNYRFDALKRVMASLASFGN 467
E            AGFINKRFDGVLAS----ELADPQYKFTDAAEVI GEAWESREFGKAVREIMALADLAN 449
          : : : : : : : . . * * : . *

Aquifex      KYVDEKQPWALNK-ERKKEELQKVLYALVDGLFVLTLLYPIT---PNKMKALQMLG-- 467
Clostridium  KYIDETMPWALAKDETKKGELDTVLYNLCE SIRI IATLINPIMNETANKIYEHIGKGGQD 481
Thermus      RYIN EKKPWELFK--KEPEEARAVLYRVVEGLRIASILLTPAM---PDKMAELRRALG-- 468
pyrococcus   RYFDHKQPWKTAKED--KVRTGTIVNISLQIVKALGILLEPFLPDASEKIWHLLNLDEVK 525
E            RYVDEQAPVWVAKQEGRDADLQAITCSMGINLFRVLMTYLKPVLKLT ERAEAFLN-TELT 508
          : * : . * * * : . : * : : :
    
```

PSIPRED) as inputs. The GenTHREADER method also runs PSI-BLAST on the target sequence before the search. *A. aeolicus* (45% identity) and *T. thermophilus* (38% identity) are likely the closest proteins structurally (Table 3).

Template analysis therefore suggested *A. aeolicus* MetRS as the main template, however the *A. aeolicus* MetRS crystal structure (2CSX) lacks the zinc binding domain involving residues 124 to 158. Employing the missing sequence folds from *T. thermophilus* MetRS rectified this problem. This multi-template approach is quite appropriate here because the missing residues are found clustering in the structure just around the zinc. Visualization of both templates in MOE by superimposition has shown that the *T. thermophilus* (1A8H) and *A. aeolicus* (2CSX) MetRS structures overlap well with the zinc binding domain of 1A8H clearly identified (Fig. 3).

Both the *T. thermophilus* and *A. aeolicus* MetRS sequences are too short to align with the *C. difficile* MetRS sequence C terminus. The C terminal structure of *P. Abyssii* MetRS (1MKH) seems to be the closest in terms of sequence identity and secondary structure; however a considerable gap in alignment of the C terminal sequence would prevent using it as a template for modeling the C terminus alone. It has been observed that many crystal structures of MetRS have an incomplete C terminal structure. For example, the structure of *P. Abyssii* MetRS (1RQG) lacks the whole C-terminus. The C-terminal domain is less well conserved overall among the MetRS enzymes, which makes modeling this domain challenging [4, 31]. Studies of the C-terminal domains of *P. Abyssii* MetRS [32] and *E. coli* MetRS [33] have indicated that the enhancement of tRNA-binding affinity conferred by the C-



Fig. 3 Superimposition of 2CSX (purple) and 1A8H (green)

terminal domain is only expressed when the enzyme is in the dimeric state and, for *E. coli* MetRS, the enzyme remains active in monomeric form after removal of a C-terminal region required for dimerization [33]. By weighing the risks and assessing the need for modeling the C terminus of the query, it was decided not to incorporate an unsatisfactory C-terminus crystal structure in the

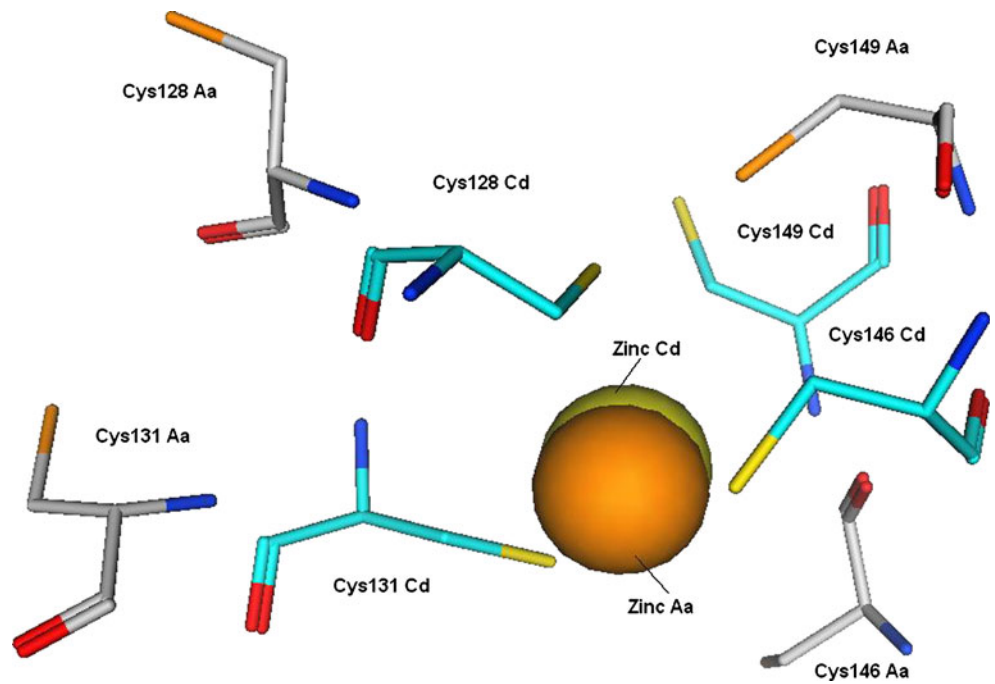
Table 3 mGenTHREADER results for *C.difficile* MetRS query

Conf	Net score	p-Value	E_{Pair}	E_{Solv}	Aln Score	Aln Len	Str Len	Seq Len	% Iden	PDB code	Description
CERT	178.758	5e-17	-650.4	-36.9	853.3	462	465	645	45.2	2ct8A0	Aquifex aeolicus MetRS
CERT	174.608	1e-16	-578.6	-27.4	857.0	504	544	645	21.0	3h9cA0	Escherichia coli MetRS
CERT	172.541	2e-16	-447.3	-26.9	867.0	545	606	645	22.4	1rqgA0	Pyrococcus abyssi MetRS
CERT	171.741	2e-16	-518.2	-39.1	825.0	490	500	645	37.8	2d5bA0	Thermus thermophilus MetRS
CERT	135.154	1e-12	-419.6	-14.0	642.0	561	862	645	17.7	1gaxA0	Thermus thermophilus ValRS
CERT	106.519	9e-10	-322.7	-15.3	469.0	521	821	645	14.0	1ileA0	Thermus thermophilus IleRS
CERT	103.727	2e-09	-256.6	-6.5	476.0	580	917	645	13.2	1ffyA0	Staphylococcus aureus IleRS
CERT	97.544	7e-09	535.5	13.1	389.0	362	399	645	9.8	3c8zA0	Mycobacterium smegmatis CysRS
CERT	91.009	3e-08	-151.6	3.2	457.0	444	628	645	9.6	2zueA0	Pyrococcus horikoshii ArgRS
CERT	90.077	4e-08	-318.0	-3.9	396.0	440	810	645	10.9	2v0cA0	Thermus thermophilus LeuRS

Conf = Description of confidence level, CERT: p-value<0.0001; HIGH: p-value < 0.001

Net Score = Raw score from SVM; p-Value = Probability of false positive; PairE = Pairwise energy for model; SolvE = Solvation energy for model; Aln Score = Sequence alignment score; Aln Len = Length of alignment; Str Len = Length of PDB entry; Seq Len = Length of target sequence; % ID = percentage identical residues between target protein and PDB protein; PDB_ID = PDB identifier (+ chain code + domain code in CATH format)

Fig. 4 Superimposition of *C. difficile* model residues (cyan, Cd) and *T. thermophilus* zinc binding template (1A8H, grey, Aa) residues interacting with the zinc ion (yellow sphere *C. difficile* model; orange sphere *T. thermophilus* template)

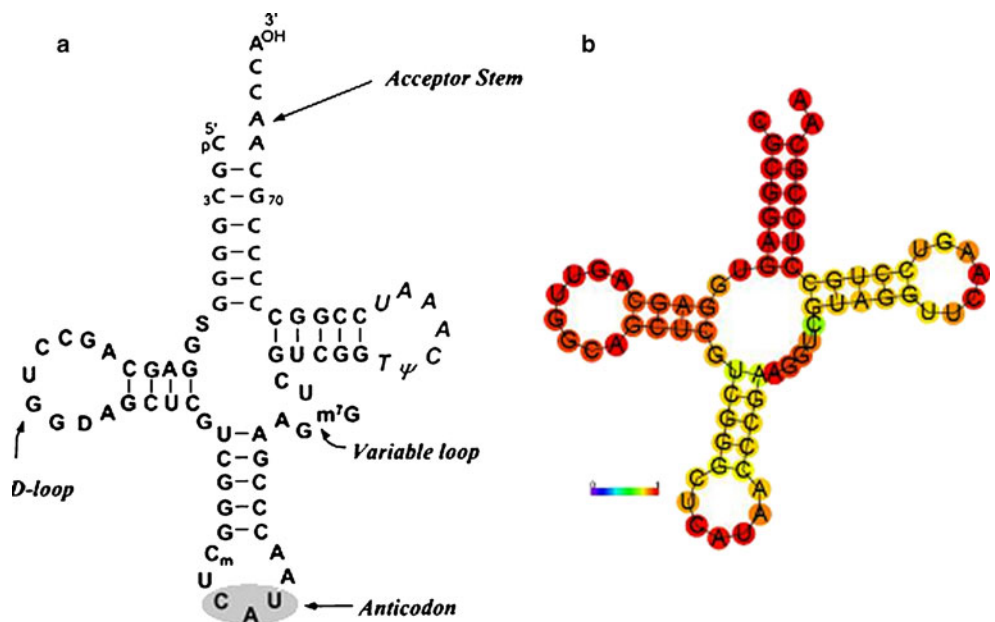


multiple template combination, which may adversely affect the rest of the model.

PSIPRED [20] secondary structure prediction from the mGenTHREADER program [22] for *C. difficile* MetRS clearly showed the high helix content predicted throughout the sequence. The C-terminus had a higher degree of coils and strands predicted than for the rest of the protein sequence, this in addition to the poor alignment observed for the C-terminus would make the C terminus a challenging sequence to model. The HIGH region motif (Table 2) is

a sequence of about ten residues and was found in most of the homologous enzymes at the sequence positions just before the 22nd residue in the N termini of these enzymes. This particular area of the query sequence was expected by PSIPRED to fold in coils and strands, this holds for *A. aeolicus* where the resolved structure shows residues 6-15 form a strand very close to where the HIGH region motif was suggested. The KMSKS motif is another known motif among the MetRS enzymes described. The motif contains the ATP binding site and has been found in the query

Fig. 5 a. *A. aeolicus* Met tRNA structure [22]. b. predicted secondary structure of the *C. difficile* Met tRNA



3D homology model

Homology models, constructed using all the *A. aeolicus* MetRS (2CSX) crystal structure and the *T. thermophilus* MetRS zinc finger residues (1A8H), were generated using Molecular Operating Environment (MOE) [23] software as described in the experimental section. Using combined templates to generate the *C. difficile* MetRS model allowed identification of the cysteine residues binding the zinc ion. These cysteine residues were all near the position of the zinc, at least three of which had the sulfur directly pointing towards the zinc. For the fourth cysteine residue side chain manipulation using MOE-rotamer explorer was employed to improve the zinc binding. The sulfur atoms binding the zinc at the constructed model were from Cys128, Cys131, Cys146 and Cys149 (Fig. 4), also noted in close proximity was Asp148 that could also chelate with the zinc ion.

The Met tRNA sequence of *C. difficile*, available from the National Centre for Biotechnology Information [34, 35] consists of 73 bases, similar to that of *A. aeolicus*. Moreover, the CAU anticodon positions are also the same. RNAfold [36], a web server for predicting the RNA secondary structure, indicated that the *A. aeolicus* Met tRNA secondary structure is quite similar to the predicted *C. difficile* Met tRNA structure using the RNA fold (Fig. 5) therefore the tRNA chain of *A. aeolicus* MetRS was used in the query MetRS homology modeling.

Asn356, Arg360 and Trp430 of *C. difficile* MetRS are fully involved in binding the anticodon part of Met-tRNA (Fig. 6). The corresponding residues in the MetRS enzymes of *A. aeolicus*, *T. thermophilus* and *E. coli* play the same role. Asn356 and Arg360 form hydrogen bonds with 35A and 34 C. As seen in the *A. aeolicus* MetRS complex the Trp residue (422 in *A. aeolicus*, Trp430 in the query sequence) is not involved in direct interaction with the anticodon 36U but interacts by hydrogen bond with a Phe residue (492 in *A. aeolicus* and Phe508 in the query sequence). While these findings are predictable by literature mining along with sequence alignments, having found that the model met these expectations can give some confidence to additional observations: Lys367 and Lys434 of *C. difficile* MetRS interact with 35A and 34 C while Leu507 and Arg510 interact with 36U (Fig. 7).

The final *C. difficile* MetRS model (Fig. 7a) contained 513 residues from the 645 residues of the sequence; this was the result of disabling the C-terminus. The secondary structure of these residues involved twenty α -helices and 14 β -sheets (Table 4). This model also contained a zinc metal and *A. aeolicus* tRNA.

The Rossmann fold domain is formed by two polypeptide segments (residues 1-115 and 229-289, colored orange in Fig. 7b). Consistent with other AaRS enzymes [29, 37, 38], the last residue of the Rossmann

fold, His289 in the β 12 strand, forms a hydrogen bond the backbone N-H group of Val259 in the β 11 strand and is followed by a glycine residue. The HIGH motif of *C. difficile* MetRS (His20-Ile21-Gly22-His23) is located between β 2 and α 1, with the two histidine residues stacked upon each other.

Inserted between the two Rossmann fold polypeptide segments is the connective polypeptide (CP) domain, composed of an antiparallel β sheet core (cyan color, Fig. 7b), a MetRS-specific β -strand insertion (green, Fig. 7b), zinc finger (yellow, Fig. 7b) and antiparallel

Table 4 Secondary structure of *C.difficile* MetRS model

Residues	2 ^o structure	Domain
PHE6-TYR9	β 1	Rossmann fold
ILE12-TYR13	β 2	
ILE21-ARG38	α 1	
VAL44-THR50	β 3	
GLU55-ILE62	α 2	
GLU68-MET86	α 3	
ASP92-ARG95	β 4	
LYS99-GLN115	α 4	
ILE118-TRY127	β 5	
GLU137-GLN139	α 5	Zinc-finger
LEU155-PHE163	β 6	Antiparallel β -sheet
GLU169-LYS177	α 6	α 2 helical structure
ARG187-LYS198	α 7	
LEU203-SER204	β 7	Antiparallel β -sheet
THR206-ARG207	β 8	
LYS215-VAL216	β 9	
VAL223-ILE224	β 10	MetRS-specific insertion
ASP229-THR236	α 8	Rossmann fold
GLU246-LYS250	α 9	
GLN257-GLY260	β 11	
LYS261-VAL264	α 10	
ILE270-ALA277	α 11	
VAL286-HIS289	β 12	
ILE292-LEU293	β 13	Stem contact (SC) fold
PRO309-ARG315	α 12	
THR320-LEU326	α 13	
GLY335-SER336	β 14	
HIS339-TYR268	α 14	α -Helix bundle
LEU386-ASP402	α 15	
PHE406-TYR423	α 16	
PRO429-LYS434	α 17	
ASP443-ILE460	α 18	
ASN465-ASP481	α 19	

α_2 -helical structure (blue, Fig. 7b). The CP core consists of β_5 , β_6 , β_7 , β_8 and β_9 , between the second and third strands (β_6 and β_7) is the zinc finger with one Zn^{2+} ion coordinated by sulfur atoms (Cys128, Cys131, Cys146 and Cys149, Fig. 4). In *T. thermophilus* MetRS the zinc binding domain consists of an antiparallel β -sheet insertion involving 3 β -strands [29]. In this *C. difficile* MetRS model, the zinc binding domain consists of a loop between β_5 and β_6 with an α -helix (α_5) positioned in the middle of the four zinc-coordinating cysteine residues. Between the second and third β -strands (β_6 and β_7), two antiparallel α -helices (α_6 and α_7) are inserted (Fig. 7b), and between the last antiparallel β -sheet domain and the C-terminal Rossmann fold domain, a β -strand (β_{10} , Val223-Ile224, green in Fig. 7b) is inserted in an antiparallel manner. This β -strand insertion is considered to be a MetRS specific element [29].

The region connecting the N-terminal Rossmann fold (orange, Fig. 7b) with the C-terminal α -helix bundle domain (magenta, Fig. 7b) is the SC-fold domain (red, Fig. 7b). The SC-fold domain, which is the RNA binding domain, has a β - α - α - β - α (β_{13} - α_{12} - α_{13} - β_{14} - α_{14}) topology. This homologous topology is well conserved and characteristic of the monomeric class Ia and Ib synthetase enzymes [29, 37, 38]. The KMSKS (Lys298-Met299-Ser300-Lys291-Ser292) motif is situated in the SC-fold domain in a loop between β_{13} and α_{12} .

The anticodon binding α -helix bundle domain (magenta, Fig. 7b) consists of five antiparallel α -helices (α_{15} - α_{19}).

The amino acid residues involved in anticodon recognition are clustered on one face of the α -helix bundle domain, with Asn356, Arg360 and Lys367 in the α_{14} helix and Trp430 and Lys434 in the α_{16} helix.

Model validation

To evaluate the quality of the modeled structures, the lowest energy model generated was subjected to a number of checks. Stereochemical quality was assessed using Ramachandran plots, using the Cambridge RAMPAGE server [25], and amino acid environment was assessed using ERRAT [26]. Overall protein structure was assessed using ProSA [27].

Validation results would suggest that the model performed well in terms of mainchain stereochemistry and amino acid environment. In the Ramachandran plot, a total of 97.9% of the residues were in the allowed region, which was comparable with the main template 2CSX (97.1%), indicating that the backbone dihedral angles ψ and φ in the model were reasonably accurate. There were 11 outliers in the model however they were far from the active site and do not significantly contribute to its function. The ERRAT results showed that the overall quality factor of the protein was good with a score of 86.139 compared with the main template 2CSX (84.862). It is surprising for the model to perform better than the template, however the template was missing the Zn-

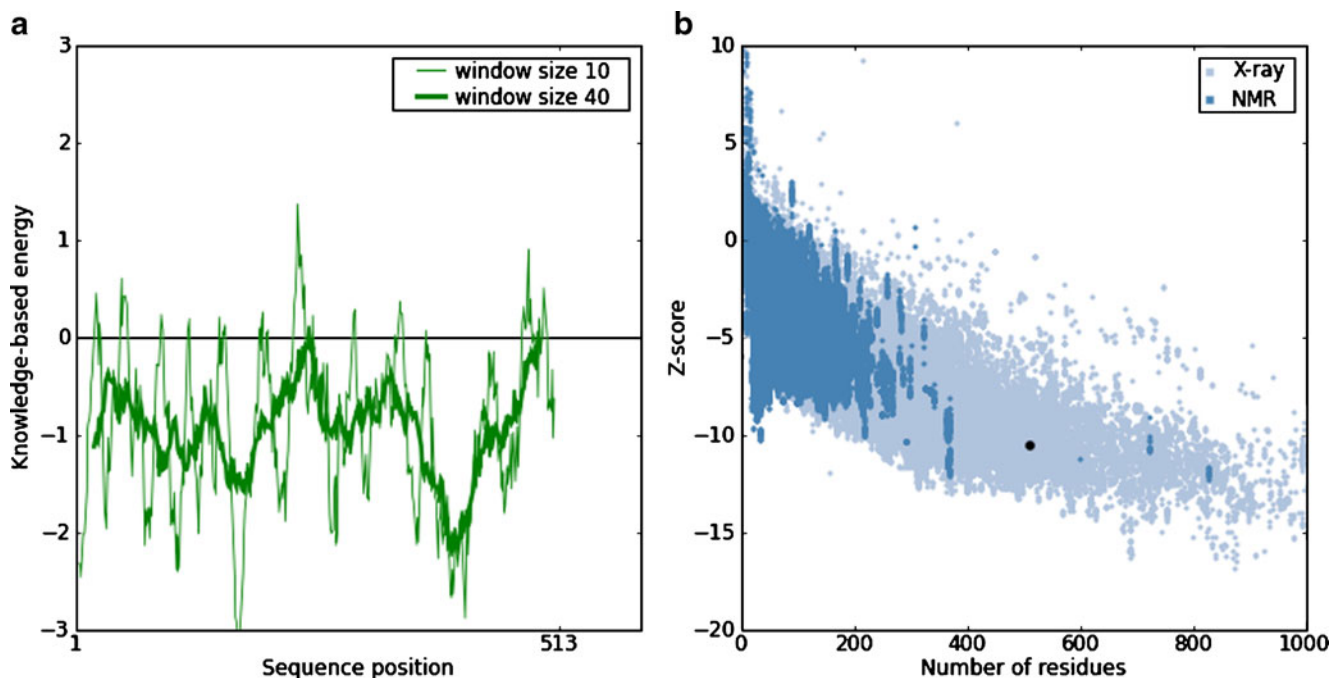


Fig. 8 ProSA output for the *C. difficile* Met tRNA model. **a.** shows the local model quality by plotting energies as a function of amino acid sequence position. **b.** indicated the z-score (dark spot) on a plot

of z-scores of all protein chains in the PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue)

Table 5 Representative dockings into *C. difficile* MetRS model

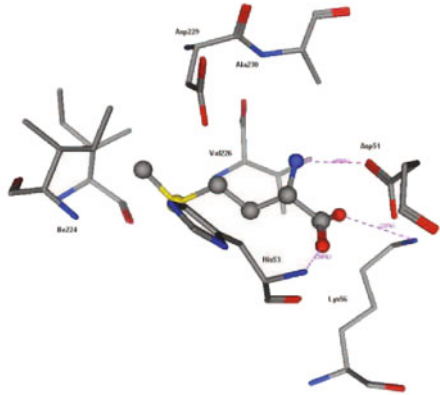
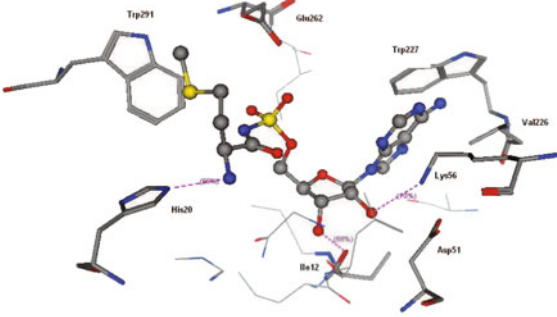
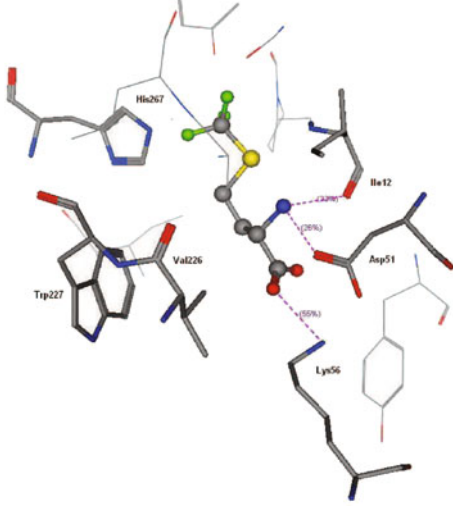
Substrate	Docking	H-bonds
Methionine		Asp51, His53, Lys56
MSP		Ile12, His20, Lys56
MF3		Ile12, Asp51, Lys56

Table 5 (continued)

2FM (pocket 1)		Asp51, Lys56
2FM (pocket 2)		Glu55, Ser133, Tyr225
MPH (pocket 1)		Asp51, Lys56
MPH (pocket 2)		Glu55, Ser133, Tyr225

binding domain residues. The Zn-binding template 1A8H scored better in both Ramachandran (99.4 % of residues in the allowed region, 3 outliers) and ERRAT (98.374) evaluation, therefore the combined template approach has incorporated some improvements in stereochemical quality compared with the 2CSX template alone.

ProSA analysis provides output in two plots: the first plot (Fig. 8a) shows the local model quality by plotting energies as a function of amino acid sequence position, in general, positive values correspond to problematic or erroneous parts of the input structure. The second plot (Fig. 8b) shows overall model quality from which the z-score is calculated, its value is displayed in a plot that contains the z-scores of all experimentally determined protein chains in the current PDB, a negative score indicates a good model whereas a positive score would indicate errors. The z-score of the model was -10.57 compared with 2CSX (-11.79) and 1A8H (-12.72). Furthermore, superimposition of the model with the main template 2CSX showed a very low RMSD (0.816 \AA) suggesting a high similarity between them, superimposition with 1A8H resulted in a higher RMSD (2.198 \AA), however as only the zinc finger residues were used from 1A8H this was not expected to have a detrimental effect on the model.

Overall based on Ramachandran plot, ERRAT and ProSA a good model, in terms of quality of the local geometry of the backbone and side-chain conformations, for *C. difficile* MetRS was generated. Further validation of the active site architecture and key binding interaction were achieved by substrate and ligand docking experiments.

Active site analysis and docking

By using the ClustalW2 multiple sequence alignment [21] and the MOE alignment service [23], the positions of the active site conserved residues in the query sequence, *A. aeolicus*, *E. coli* and *T. thermophilus* MetRS can be identified. Docking suitable ligands into the constructed model was used to validate the predicted active site. The docked ligands include the MetRS natural substrate (methionine), several *E. coli* MetRS inhibitors 5'-O-[N-(L-methionyl)-sulfamoyl] (MSP), difluoromethionine (2FM) and trifluoromethionine (MF3), (1-amino-3-methylsulfanylpropyl)-phosphonic acid (MPH) and a *C. difficile* MetRS inhibitor (REP3123), no site restriction was specified in the MOE docking procedures.

Docking the natural substrate, methionine, identified key residues with His53, Asp51 and Lys56 consistently forming H-bonding interactions with the carboxylic acid moiety of methionine. Ile12 was also found to hydrogen bond with this moiety in different dockings and the residues Ile224, Val226 and Ala230 forming a hydrophobic pocket (Table 5).

The MSP ligand was docked in the *C. difficile* MetRS model. It was observed that the residues around the docked ligand remained almost the same in all the dockings (Table 5). The following residues interact with the ligand in different docking attempts: Ile12, His20, Asp51, Lys56, Val226, Trp227 and Trp291 with H-bonding interactions most frequently observed with Asp51, Lys56 and Ile12. Difluoromethionine (2FM) and trifluoromethionine (MF3) are only available bound to the *E. coli* MetRS [39]. MOE dockings of MF3 showed that Asp51 of the *C. difficile* MetRS model is consistently interacting with the ligand. Ile12, Asp51, His267, Val226, Trp227, Leu233 and Ile263 are in the model binding pocket surrounding the ligand and are very close to the expectations derived from *E. coli* MetRS complex with MF3 (Table 5).

2FM was found to dock in two sites within the active site, the first was the site involving interaction with Asp51 and Lys56 seen previously with methionine, MSP and MF3 and the second site involved interaction with Glu55, Ser133 and Tyr225. Many conserved residues in the N terminal and in positions from 224 to 240 and from 255 to 268 are also found in the 2FM dockings (Table 5). Docking of 1-amino-3-methylsulfanylpropyl)-phosphonic acid (MPH), found in the resolved complex structure of *E. coli* MetRS (1P7P) [40] resulted in diverse ligand interactions but comparable with those found when docking 2FM, that is MPH docked in two distinct binding sites within the active site. Again the first was the site involving interaction with Asp51 and Lys56 seen previously with MSP and MF3 and the second site involved interaction with Glu55, Ser133 and Tyr225 (Table 5).

The binding mode of ligands in the *C. difficile* MetRS model was comparable with the binding of identical ligands in *E. coli* crystal structures (PDB identifier: 1F4L (methionine); 1PFY (MSP); 1PFV (2FM); 1PFW (MF3); 1P7P (MPH)). In these crystal structures the key H-bonding

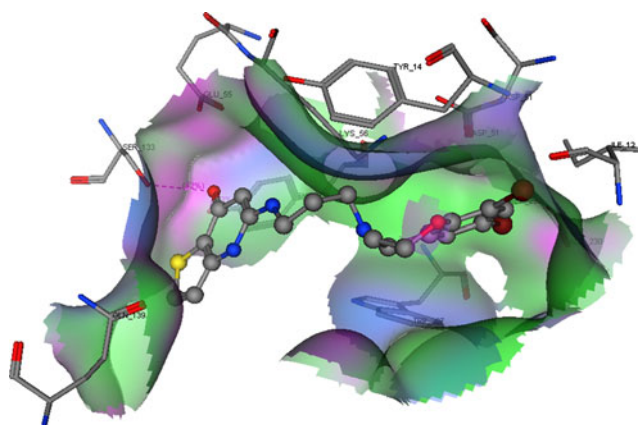


Fig. 9 Representative REP3123 docking into *C. difficile* MetRS model showing Gaussian molecular surface contacts. H-bonding (magenta), hydrophobic (green), mild polar (blue)

interactions involved the conserved residues Leu13 and Asp52, with the residues Ala256 (Ala230 in *C.difficile*), Pro257, Tyr260, Ile297 (Ile263 in *C.difficile*) and His301 (His267 in *C. difficile*) forming hydrophobic interactions. The overlap in binding residues between identical ligands and the *C. difficile* MetRS model and *E. coli* MetRS crystal structures demonstrates that the modeled structure is of sufficient quality to correctly reproduce ligand binding poses in docking simulations.

REP 3123 is the only ligand found in the literature that has been described as a selective *C. difficile* MetRS inhibitor [13]. The antimicrobial effect is described and the chemical structure of the inhibitor. It was confirmed in their study that REP 3123 is a selective and potent competitive inhibitor of *C. difficile* MetRS, which implies that the ligand has features that enable it to strongly bind at the active site of *C. difficile* MetRS. REP 3123 is not a competitive inhibitor of ATP and the presence of ATP facilitated inhibition [13], therefore it could be deduced that the ligand binds by a similar mechanism as methionine. REP 3123 does not inhibit *E. coli* MetRS however it does inhibit *S. aureus* MetRS [13], which may be explained by the close position of *C. difficile* and *S. aureus* MetRS indicated by the phylogenetic tree (Fig. 1).

Docking of REP3123 indicated that this ligand docks across both binding sites, the first pocket is formed by Asp51, Ile12, Ala230 and Trp227 and the second pocket, linked by Lys56 and Tyr14, is formed by Glu55, Ser133, Tyr225 and Gln139 (Fig. 9). Interestingly a pharmacophore virtual screening study by Finn et al [10], employed to develop *S. aureus* MetRS inhibitors, identified Asp51 as a key residue with optimal inhibitory activity observed on formation of two hydrogen bond interactions between the ligand and the critical Asp51 group. Another study investigating the effect of resistance mutations on *S. aureus* MetRS activity found that Gly54 (G54S) and Gly223 (G223C) mutations were the most deleterious single mutants with markedly impaired enzymatic function [41]. The high sequence identity (50%) between *S. aureus* MetRS and *C. difficile* MetRS may suggest that the proximity of Glu55 to Gly54 in the binding pocket would indicate the importance of Glu55 in binding interactions and inhibition of enzyme function. The ability of REP3123 to interact with both binding pockets and form H-bonding interactions with several residues including Asp51, Ile12, Lys56, Glu55, Ser133 and Tyr225 may account for the high selectivity and inhibitory activity against *C. difficile* MetRS.

Conclusions

Development of a homology model for MetRS has been explored using a multi-template approach and the generated

C.difficile MetRS model has the characteristic domains and topology found in class I AaRS enzymes. Docking suitable ligands into the constructed model was then used to validate the predicted active site. The docked ligands included the MetRS natural substrate (methionine), several *E. coli* MetRS inhibitors and a *C. difficile* MetRS inhibitor (REP3123), no site restriction was specified in the MOE docking procedures. Both conservation analysis and docking studies have agreed to the same active site with two pockets identified, the active site key residues in the model, which span the C-terminal Rossmann fold and the CP domain, include Ile12, Asp51, Ala230, Trp227 (pocket 1), Glu55, Ser133 and Tyr225 (pocket 2) with Lys56 bridging the two pockets. The ability to interact with both pockets within the binding site with multiple hydrogen bonding interactions, as observed for REP3123, may be advantageous for inhibitory activity. The Met-tRNA binding within the constructed model has also been validated. In the constructed model, Asn356, Arg360 and Trp430 were conserved and involved in the specific interactions between the protein and the Met-tRNA anticodon (CAU), the corresponding residues in close homologous MetRS have the same binding role. Having identified both the active site pockets of *C. difficile* MetRS and key binding interactions, this model can serve as a tool for drug design and the development of inhibitors with greater selectivity can be explored.

References

1. Richmond MA, Riggs MM, Eckstein BC, Donskey CJ (2008) Clostridium difficile infection in patients with SCI. J Spinal Cord Med 31:521–521
2. Gerding DN, Muto CA, Owens RC (2008) Treatment of Clostridium difficile infection. Clin Infect Dis 15:S32–42
3. Deaths involving Clostridium difficile: England and Wales 2004–2008 Office for National Statistics 19/8/2009. <http://www.statistics.gov.uk>
4. Deniziak MA, Barciszewski J (2001) Methionyl-tRNA synthetase. Acta Biochim Pol 48:337–350
5. Hountondji C, Dessen P, Blanquet S (1986) Sequence similarities among the family of aminoacyl-tRNA synthetases. Biochimie 68:1071–1078
6. Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of transfer-RNA synthetases into 2 classes based on mutually exclusive sets of sequence motifs. Nature 347:203–206
7. Hurdle JG, O'Neill AJ, Chopra I (2005) Prospects for aminoacyl-tRNA synthetase inhibitors as new antimicrobial agents. Antimicrob Agents Chemother 49:4821–4833
8. Boyce JM (2001) MRSA patients: proven methods to treat colonization and infection. J Hosp Infect 48:S9–14
9. Jarvest RL, Berge JM, Berry V, Boyd HF, Brown MJ, Elder JS, Forrest AK, Fosberry AP, Gentry DR, Hibbs MJ, Jaworski DD, O'Hanlon PJ, Pope AJ, Rittenhouse S, Sheppard RJ, Slater-Radosti C, Worby A (2002) Nanomolar inhibitors of Staphylococcus aureus methionyl tRNA synthetase with potent antibacterial activity against gram-positive pathogens. J Med Chem 45:1959–1962

10. Finn J, Mattia K, Morytko M, Ram S, Yang Y, Wu X, Silverman J, Mak E, Gallant P, Keith D (2003) Discovery of a potent and selective series of pyrazole bacterial methionyl-tRNA synthetase inhibitors. *Bioorg Med Chem Lett* 13:2231–2234
11. Tandon M, Coffen D, Gallant P, Keith D, Ashwell M (2004) Potent and selective inhibitors of bacterial methionyl tRNA synthetase derived from an oxazolone-dipeptide scaffold. *Bioorg Med Chem Lett* 14:1909–1911
12. Finn J, Stidham M, Hilgers M, Kedar GC (2008) Identification of novel inhibitors of methionyl-tRNA synthetase (MetRS) by virtual screening. *Bioorg Med Chem Lett* 18:3932–3937
13. Critchley IA, Green LS, Young CL, Bullard JM, Evans RJ, Price M, Jarvis TC, Guilles JW, Janjic N, Ochsner U (2009) Spectrum of activity and mod of action of REP3123, a new antibiotic to treat *Clostridium difficile*. *Antimicrob Agents Chemother* 63:954–963
14. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784–3788. www.expasy.org
15. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* 38:779–786
16. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
17. RCSB Protein Data Bank (PDB) <http://www.rcsb.org/pdb>
18. <http://www.phylogeny.fr/>
19. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:W465–469
20. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202
21. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. *Bioinformatics* 23:2947–2948
22. Lobley A, Sadowski MI, Jones DT (2009) pGenTHREADER and pDomTHREADER: New Methods For Improved Protein Fold Recognition and Superfamily Discrimination. *Bioinformatics* 25:1761–1767
23. Molecular Operating Environment (MOE 2008.09) Chemical Computing Group Inc, Montreal Quebec Canada <http://www.chemcomp.com>. 2008.09
24. Weiner SJ, Kollman PA, Nguyen DT (1986) An all atom forcefield for simulations of proteins and nucleic acids. *J Comput Chem* 7:230–252
25. RAMPAGE Server <http://ravenbioccam.ac.uk/rampage.php>
26. UCLA-DOE Institute for Genomics & Proteomics Server <http://www.doe-mbiucla.edu/Services>
27. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410
28. Landes C, Perona JJ, Brunie S, Rould MA, Zelwar C, Steitz TA, Risler JL (1995) A structure-based multiple sequence alignment of all class I aminoacyl-tRNA synthetases. *Biochimie* 77:194–203
29. Sugiura I, Nureki O, Ugaji-Yoshikawa Y, Kuwabara S, Shimada A, Tateno M, Lorber B, Giege R, Moras D, Yokoyama S, Konno M (2000) *Structure* 8:197–208
30. Kamijo S, Fujii A, Onodera K, Wakabayashi K (2009) Analysis of conditions for KMSS loop in Tyrosyl-tRNA synthetase by building a mutant library. *J Biochem* 146:241–250
31. Spencer AC, Heck A, Takeuchi N, Watanabe K, Spremulli LL (2004) Characterization of the human mitochondrial methionyl-tRNA synthetase. *Biochemistry* 43:9743–9754
32. Crepin T, Schmitt E, Blanquet S, Mechulam Y (2004) Three-dimensional structure of methionyl-tRNA synthetase from *Pyrococcus abyssi*. *Biochemistry* 3:2635–2644
33. Mechulam Y, Schmitt E, Maveyraud L, Zelwer C, Nureki O, Yokoyama S, Konno M, Blanquet S (1999) Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features. *J Mol Biol* 294:1287–1297
34. Xu B, Krudy GA, Rosevear PR (1993) Identification of the metal ligands and characterization of a putative zinc finger in methionyl-tRNA synthetase. *J Biol Chem* 268:16259–16264
35. RNAfold server: <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>
36. Mathews DH, Sabina J, Zuker M, Turner H (1999) Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J Mol Biol* 288:911–940
37. Cavarelli J, Delagoutte B, Eriani G, Gangloff J, Moras D (1998) L-arginine recognition by yeast arginyl-tRNA synthetase. *EMBO J* 17:5438–5448
38. Nureki O, Vassylyev DG, Katayanagi K, Shimizu T, Sekine S, Kigawa T, Miyazawa T, Yokoyama S, Morikawa K (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetases. *Science* 267:1958–1965
39. Morales AJ, Swairjo MA, Schimmel P (1999) Structure-specific tRNA-binding protein from the extreme thermophile *Aquifex aeolicus*. *EMBO J* 18:3475–3483
40. Crepin T, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF, Blanquet S (2003) Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase. *J Mol Biol* 332:59–72
41. Green LS, Bullard JM, Ribble W, Dean F, Ayers DF, Ochsner UA, Janjic N, Jarvis TC (2009) Inhibition of methionyl-tRNA synthetase by REP8839 and effects of resistance mutations on enzyme activity. *Antimicrob Agents Chemother* 53:86–94